nature microbiology

ANALYSIS

https://doi.org/10.1038/s41564-020-0777-y

Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems

Adi Millman, Sarah Melamed, Gil Amitai and Rotem Sorek ✉

Cyclic-oligonucleotide-based anti-phage signalling systems (CBASS) are a family of defence systems against bacteriophages (hereafter phages) that share ancestry with the cGAS–STING innate immune pathway in animals. CBASS systems are composed of an oligonucleotide cyclase, which generates signalling cyclic oligonucleotides in response to phage infection, and an effector that is activated by the cyclic oligonucleotides and promotes cell death. Cell death occurs before phage replication is completed, therefore preventing the spread of phages to nearby cells. Here, we analysed 38,000 bacterial and archaeal genomes and identified more than 5,000 CBASS systems, which have diverse architectures with multiple signalling molecules, effectors and ancillary genes. We propose a classification system for CBASS that groups systems according to their operon organization, signalling molecules and effector function. Four major CBASS types were identified, sharing at least six effector subtypes that promote cell death by membrane impairment, DNA degradation or other means. We observed evidence of extensive gain and loss of CBASS systems, as well as shuffling of effector genes between systems. We expect that our classification and nomenclature scheme will guide future research in the developing CBASS field.

Bacteria and archaea have active defence systems to protect themselves against infections by viruses[1–4]. Recent studies in *Vibrio cholerae* and *Escherichia coli* have reported a bacterial defence system that is thought to share ancient common ancestry with the animal cGAS–STING antiviral pathway[5]. The bacterial cyclic GMP-AMP synthase (cGAS)[6] generates cyclic GMP-AMP (cGAMP) molecules after sensing that the cell has been infected, and the cGAMP molecules activate a phospholipase[7] that degrades the inner membrane of the bacterial cell, resulting in cell death[5]. cGAS-mediated cell death occurs before the phage is able to complete its replication cycle, such that no mature phage particles emerge from the infected cell and the phage does not spread to nearby cells[5]. This mode of bacterial defence was named CBASS[5]. In general, CBASS systems are characterized by at least two proteins with the following minimal configuration: one protein is thought to sense the presence of the phage and then produce the cyclic-oligonucleotide signal, and the second is an effector protein that senses the cyclic-oligonucleotide signal and exerts the cell-suicide function (Fig. 1a).

Bioinformatics and functional analyses have revealed that variants of the CBASS system are widespread in bacterial and archaeal genomes, forming a large and highly diverse family of anti-phage defence systems[5,8–11]. Bacterial CBASS systems show variety in every part of the system, including the oligonucleotide cyclase, the signalling molecule produced, the effector cell-killing gene and ancillary genes[5,8–10].

Here we systematically analysed CBASS systems in microbial genomes and identified four major types of CBASS. We analysed the different protein domains that are associated with cyclic-oligonucleotide recognition, effector activity and ancillary functions, and propose a standardized nomenclature to describe CBASS diversity.

Results

The presence of an oligonucleotide cyclase gene of the CD-NTase family is the hallmark of CBASS systems[5,8]. To study the distribution and variation of CBASS systems in bacteria and archaea, we first searched for homologues of oligonucleotide cyclase genes in a set of 38,167 bacterial and archaeal genomes belonging to 14,566 species (see Methods). We found 5,756 such genes in 4,894 (13%) of the genomes, spanning—in agreement with previous analyses[5,8]—all major bacterial phyla as well as one archaeal phylum (Supplementary Tables 1–5). We next analysed the genomic environment of the oligonucleotide cyclases and identified, in the majority of cases, the effector gene and additional commonly associated ancillary genes, which are denoted CD-NTase associated proteins (Cap) genes[12] (Fig. 1b,c; see Methods). On the basis of the analysis of these identified CBASS systems, we propose a CBASS classification scheme that is based on three axes: the CBASS operon composition (CBASS types I, II, III and IV); the effector activity; and the signalling molecule produced by the oligonucleotide cyclase (Fig. 1b). Overall, we detected 5,675 cyclase-containing CBASS operons; some operons contain more than one cyclase (see below).

Classification axis I: CBASS operon configuration. *Type I CBASS comprises a compact two-gene system.* In 2,376 of the CBASS systems that we identified (42%), the system has a minimal configuration that consists of an oligonucleotide cyclase and an effector gene, without associated ancillary genes (Supplementary Table 1). As this is the most abundant CBASS configuration, we denote it type I. A type I CBASS from *Bacillus cereus*, when cloned into *Bacillus subtilis*, was demonstrated experimentally to confer protection against phages, indicating that this two-gene minimal configuration does not necessitate ancillary genes for its anti-phage activity[5]. Type I CBASS can be found in all major phyla (Fig. 2). The effector proteins in the majority of type I CBASS systems contain one or several transmembrane-spanning helices, which are predicted to form pores in the membrane once activated by the cyclic-oligonucleotide signal[5,11] (Fig. 2d).

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. ✉e-mail: rotem.sorek@weizmann.ac.il

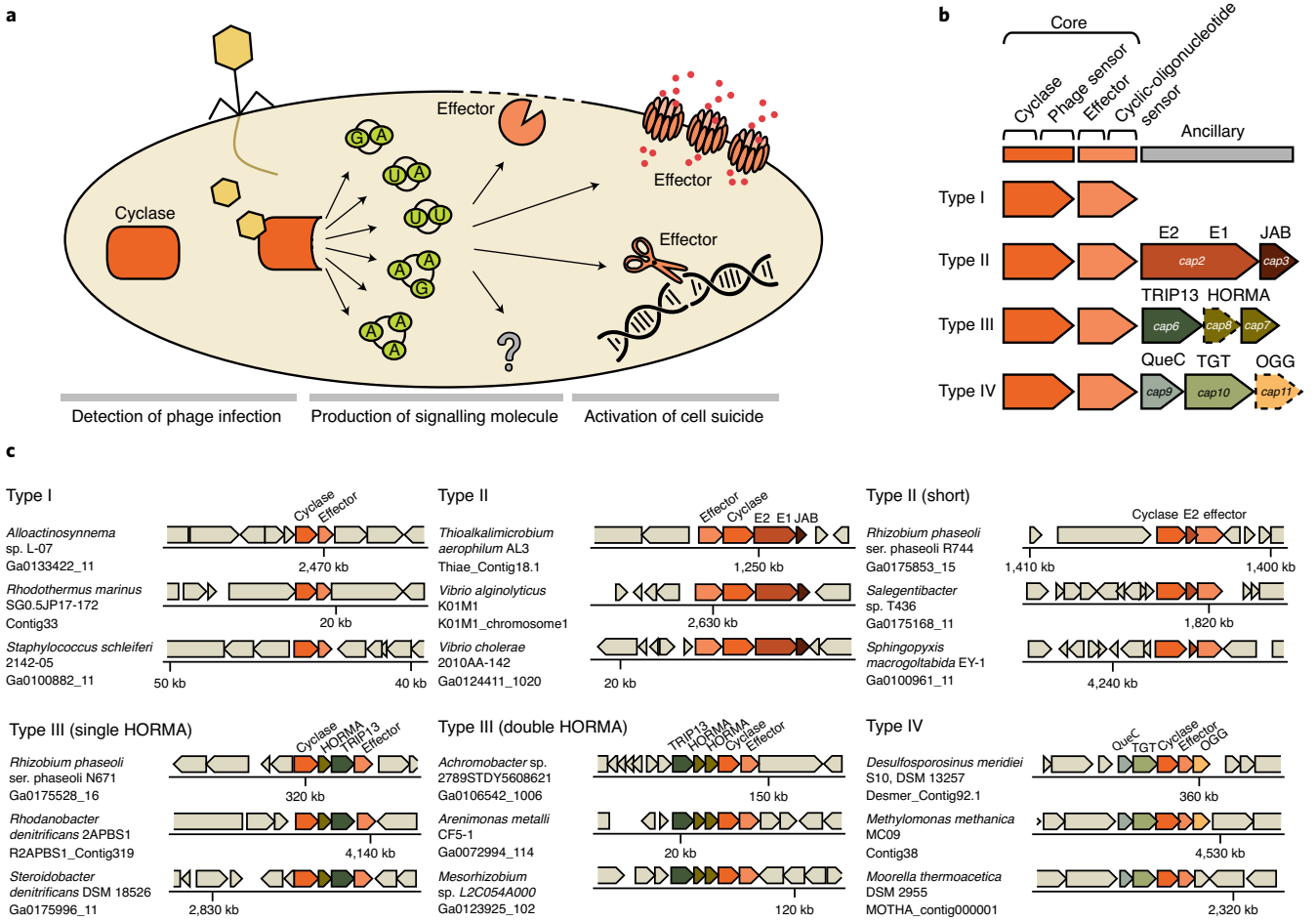


Fig. 1 | General description of CBASS systems. a, A general model for the mode of action of CBASS systems. Phage infection is sensed by the cyclase (or by the ancillary genes, together with the cyclase), leading to activation of the cyclase, which generates a cyclic-oligonucleotide signalling molecule. The signalling molecule is sensed by the effector and activates its cell-killing function. CBASS cyclases can generate a variety of cyclic di- and trinucleotides, and CBASS effectors can exert cell death through membrane degradation, cleavage of phage and host DNA, formation of membrane-spanning pores and by other means of which the mechanisms are yet to be identified. **b**, Four types of CBASS systems classified according to their ancillary gene content. The box arrows with dashed outlines represent genes that are present in some, but not all, of the systems in the respective CBASS type. Ancillary genes are denoted as Cap genes. **c**, Representative examples of CBASS operons and their genomic neighbourhood. The name of the bacterial species and the accession of the relevant genomic scaffold in the IMG database[34] are indicated on the left.

Type II CBASS includes ancillary genes that encode ubiquitin-associated domains. We found that 2,199 CBASS systems (39%) include, in addition to the core cyclase-effector pair, two genes (called *cap2* and *cap3* (ref. [12])) that encode domains that are typical of the eukaryotic ubiquitin machinery. These systems are denoted type II CBASS. *cap2* comprises a fusion between a ubiquitin-activating E1-like enzymatic domain and a ubiquitin-conjugating E2-like domain. The second gene, *cap3*, encodes a protein domain that is predicted to be an isopeptidase of the JAB/JAMM family, which is usually found in eukaryotic deubiquitinase enzymes that remove ubiquitin from target proteins. Two type II CBASS systems, one from *E. coli* and one from *V. cholerae*, were shown experimentally to confer defence against a phylogenetically wide set of coliphages, when cloned to an *E. coli* strain that lack a CBASS system[5]. Interestingly, *cap2* and *cap3* were necessary for protection against some, but not all, of the phages, suggesting that these genes have an ancillary function in the CBASS system that expands the range of phages against which the system protects. As ubiquitin is not known to be present in *E. coli* or *V. cholerae*, evidence is lacking about which protein is the target of the ubiquitin-handling domains encoded by the *cap2* and *cap3* ancillary genes. In the experimentally tested type II CBASS systems from *E. coli* and *V. cholerae*, the effector gene is a phospholipase that is thought to degrade the inner bacterial membrane once activated by the signalling molecule[5,7]; such phospholipase domains are the most common effector domains in type II CBASS systems (Fig. 2d).

As the *cap3* gene is short (encoding 160 amino acids on average), it is sometimes not annotated in the genome in which the CBASS is identified. Nevertheless, examination of the intergenic regions downstream of *cap2* enabled the identification of *cap3* in many cases (Supplementary Table 2). In a minority (19) of type II CBASS systems, *cap2* and *cap3* are fused and, in other cases (13), *cap3* is fused to the effector gene. A subset of 616 (28%) of the type II CBASS systems have a minimal configuration that comprises, in addition to the cyclase-effector pair, only a short E2-domain gene, without E1 or JAB (Fig. 1c and Supplementary Table 2). It is unclear whether this subtype functions on its own or whether it uses other genes in the cell to fill in for the missing E1 and JAB functions.

Type III CBASS includes ancillary genes that encode HORMA and TRIP13 domains. About 10% of the CBASS systems that we

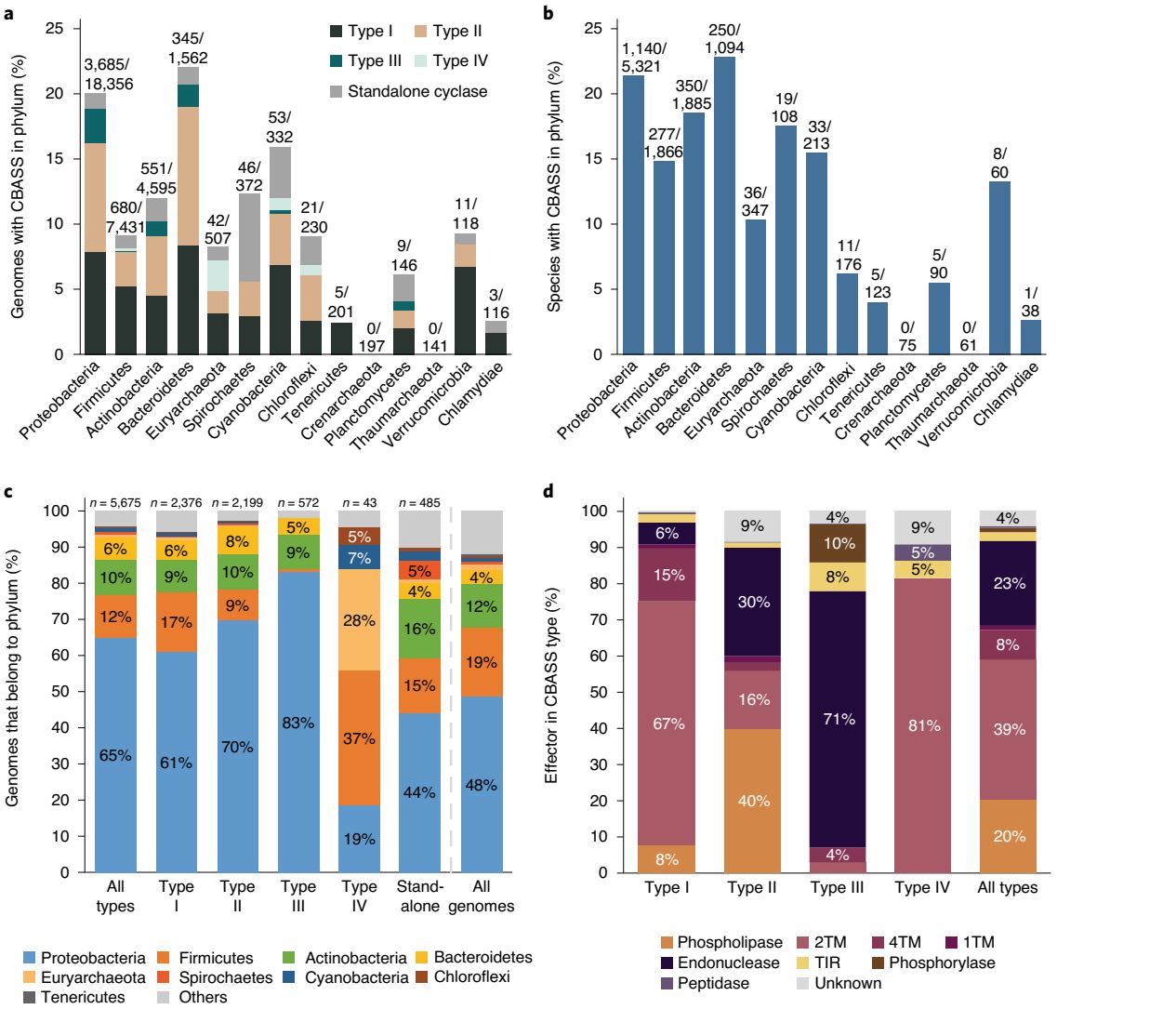


Fig. 2 | Phylogenetic distribution of CBASS types and effectors. a, The presence of CBASS systems in analysed genomes, divided by phyla. Data for phyla with >100 genomes in the database are shown. The values above each bar represent the number of genomes from the specific phylum that contain CBASS out of the total number of genomes from the specific phylum that are present in the analysed database. **b**, The presence of CBASS systems in species, divided by phyla. A CBASS system was counted as present in a species if at least one genome of that species contained a CBASS. The values above each bar represent the number of species from the specific phylum that contain a CBASS out of the total number of species from the specific phylum that are present in the analysed database. **c**, Phyletic distribution of genomes per CBASS type. Data are shown for phyla with >200 genomes in the database. The bar on the right shows the phyletic distribution of all 38,167 genomes in the database for comparison purposes. **d**, The distribution of effector genes in each CBASS type.

identified (572 systems) are associated with another set of ancillary genes that, interestingly, also encode protein domains that were to date mainly described in eukaryotes[13]. One of these genes (hereafter, *cap7*) encodes a HORMA-domain protein; such proteins form signalling complexes that control steps in meiosis, mitosis and DNA repair in eukaryotes[14]. The second gene (*cap6*) encodes a TRIP13 domain (also called Pch2 domain), which is a known inhibitor of HORMA-domain protein activity. In eukaryotes, TRIP13 proteins dissociate the HORMA signalling complexes and, therefore, turn off the signal[15].

A type III CBASS from *E. coli* MS115-1 was shown to provide protection against phage-lambda infection[9]. The oligonucleotide cyclase in this system produces cyclic triadenylate molecules (cAAA) in response to phage infection, and this molecule was shown to activate the effector protein—an endonuclease that non-discriminately and completely degrades both phage and host DNA leading to abortive infection and cell death[9,10]. It was shown that the HORMA-domain protein (Cap7) is essential for phage defence, and that the oligonucleotide cyclase becomes active only when physically bound by the HORMA-domain protein. By contrast, the TRIP13 protein Cap6 was shown to dissociate the HORMA-domain protein from the oligonucleotide cyclase, indicating that TRIP13 is a negative regulator of cyclic-oligonucleotide production[9]. On the basis of these results, it was suggested that, under normal conditions, Cap6 prevents Cap7 from associating with and activating the oligonucleotide cyclase; however, during phage infection, Cap7 recognizes the infection (perhaps by binding to a phage-encoded protein) and changes its conformation to remain bound to the oligonucleotide cyclase. This binding activates the production of the cyclic-oligonucleotide molecule, which in turn activates the effector protein leading to cell death before the phage has completed its replication cycle[9,10].

Type III systems are overrepresented in Proteobacteria and are almost completely absent from Firmicutes (Fig. 2c). These CBASS systems appear in two main configurations: one with a single HORMA-domain protein (Cap7), and the second with two HORMA-domain proteins in which one of the two HORMA-domain proteins (which we denote Cap8) is considerably divergent and is identifiable only through structure-based comparisons, but not sequence-based comparisons[9] (Fig. 1c). It was found that this divergent Cap8 HORMA domain does not activate the cyclase in vitro and it was therefore suggested that it functions as a scaffold[9]. The oligonucleotide cyclases that are present in the two type III subtypes form two distinct clades on the phylogenetic tree of the oligonucleotide cyclase family (Extended Data Fig. 1). A specific subclade of the type III CBASS systems (102 cases, 18% of type III systems) contains an additional ancillary gene with a predicted 3′–5′ exonuclease domain (Supplementary Table 3).

Type IV CBASS includes ancillary proteins with nucleotide-modifying domains. Type IV is a rare form of CBASS, with only 43 occurrences in the analysed genomes, appearing mostly in archaea and Firmicutes. Type IV CBASS comprises two ancillary genes (*cap9* and *cap10*) with protein domains that are involved in nucleotide modification. *cap9* encodes a predicted QueC enzymatic domain, which is known to convert the modified base 7-carboxy-7-deazaguanine (CDG) to 7-cyano-7-deazaguanine ($preQ_0$)[16,17]. *cap10* encodes a predicted enzyme called TGT that catalyses base exchange of a specific guanine residue with $preQ_0$ in tRNA molecules[18]. Proteins with these enzymatic domains were also shown to modify guanine residues on DNA as part of a restriction-modification system called dpd[19]. In 12 (28%) of the type IV CBASS systems, another gene (*cap11*) that is annotated as an N-glycosylase/DNA lyase (OGG) is also present in the CBASS operon (Supplementary Table 4). N-glycosylase/DNA lyases are known to remove damaged guanine bases (8-oxoG) from the DNA and to nick DNA in apurinic/apyrimidinic sites[20]. To date, no type IV CBASS systems were demonstrated to defend against phages, and the function of their ancillary genes remains obscure.

Type IV systems are enriched in archaea and form two clades on the phylogenetic tree of the oligonucleotide cyclase family (Extended Data Fig. 1), one of which is a distinct, previously unreported clade. Notably, the relatively small number of sequenced archaeal genomes in the databases (representing only 2.5% of the set of >38,000 genomes that we analysed) may have led to an under-representation of type IV CBASS systems in our set of genomes.

Standalone oligonucleotide cyclases. We detected 485 (9%) cases in which the oligonucleotide cyclase gene was not associated with effector and/or ancillary genes (Supplementary Table 5). In many cases, these standalone cyclases appear at, or near to, the edge of partially assembled genomic scaffolds, most probably representing cases in which the remainder of the CBASS operon was in the not-yet-assembled part of the draft genome. Other cases of standalone cyclases may represent degenerated, pseudogenized CBASS systems in which some of the genes were mutated or deleted, or cases in which the effector gene is not adjacent to the cyclase but is rather found elsewhere on the genome. We cannot rule out the possibility that some oligonucleotide cyclases exert their function as standalone genes. Indeed, we detected three cases in which CBASS systems appear on a single gene, comprising a fusion between the oligonucleotide cyclase and the effector gene (Supplementary Table 1).

Classification axis II: the effector cell-killing domain. The effector gene of CBASS systems is usually composed of two domains—the cyclic-oligonucleotide-sensing domain and the cell-killing domain that becomes activated once the cyclic oligonucleotide is sensed. The various cell-killing domains can be associated with several CBASS types. In our CBASS classification scheme, we mark cell-killing domains by capital letters (for example, phospholipase domain (A), transmembrane domains (B), endonuclease domain (C) and so on; see below). Thus, CBASS type I-A is a type I system with a phospholipase effector, and CBASS type II-C is a type II system with an endonuclease effector.

Table 1 | Pfams that are commonly associated with CBASS systems in prokaryotes

Pfam	Function in CBASS	Additional names
PF18144	Oligonucleotide cyclase	CD-NTase[8]
PF18134	Domain of unknown function fused to cyclases	AGS-C[11]
PF18145	Oligonucleotide-sensing domain	SAVED[11]
PF18178	Oligonucleotide-sensing domain	TPALS[11]
PF15009	Oligonucleotide-sensing domain (STING)	STING
PF18153	Effector domain (2TM)	S_2TMBeta[11]
PF18179	Effector domain (2TM)	SUa-2TM[11]
PF18303	Effector domain (2TM)	SAF 2TM[11]
PF18181	Effector domain (2TM)	SLATT[11]
PF18183	Effector domain (2TM)	SLATT[11]
PF18184	Effector domain (2TM)	SLATT[11]
PF18186	Effector domain (2TM)	SLATT[11]
PF18160	Effector domain (2TM)	SLATT[11]
PF18169	Effector domain (2TM)	SLATT[11]
PF18167	Effector domain (2TM)	SMODS-associated NUDIX domain[11]
PF18159	Effector domain (4TM)	S-4TM[11]
COG3105	Effector domain (1TM)	DUF1043
PF00899	Ancillary gene, type II (E1 domain)	
PF14461	Ancillary gene, type II (E2 domain)	
PF14464	Ancillary gene, type II (JAB domain)	
PF18173	Ancillary gene, type III (HORMA domain)	
PF18138	Ancillary gene, type III (HORMA domain)	

The following cell-killing domains are identified in CBASS effector genes. In some cases, they were shown to cause cell death; in other cases, the cell-killing effect is hypothesized but has not yet been examined experimentally.

Patatin-like phospholipase (A). Effector proteins with a patatin-like phospholipase domain were shown to degrade the phospholipids of the inner cell membrane, as demonstrated in vitro as well as by in vivo studies in *V. cholerae*[7]. This leads to a change in the cell shape and eventual cell lysis[5]. Effector genes with phospholipase domains are found both in type I and type II CBASS systems, but are absent from types III and IV. These effectors are the most common effectors in type II CBASS (Fig. 2d).

Effectors with two or four transmembrane helices (B). This effector class is composed of proteins that do not have an identifiable enzymatic domain but, rather, encode transmembrane helices. These effectors are predicted to promote cell death by forming pores in the cell membrane once they become activated by the cyclic oligonucleotide[5,11]. Effectors with two transmembrane helices (2TM) are the

Table 2 | Experimentally studied CBASS systems and their classifications

Species	Cyclase	Effector	Ancillary genes	Signalling molecule	CBASS classification	Ref.
E. coli, *V. cholerae*	DncV (cGAS)	Phospholipase	Cap2, Cap3	cGAMP	Type II-A^{GA}	5
P. aeruginosa, *Xanthomonas citri*	CD-NTase002 CD-NTase003	Phospholipase	Cap2, Cap3	cGAMP	Type II-A^{GA}	8
E. coli	CdnE	Phospholipase	Cap2, Cap3	Cyclic UMP-AMP	Type II-A^{UA}	8
Legionella pneumophila	Lp-CdnE02	2TM	None	Cyclic UMP-UMP	Type I-B^{UU}	8
Enterobacter cloacae	Ec-CdnD02	Endonuclease	Cap2, Cap3	Cyclic AMP-AMP-GMP	Type II-C^{AAG}	8,12
E. coli	CdnC	Endonuclease	Cap6, Cap7	Cyclic AMP-AMP-AMP	Type III-C^{AAA}	9
P. aeruginosa	CdnD	Endonuclease	Cap6, Cap7, Cap8	Cyclic AMP-AMP-AMP	Type III-C^{AAA}	9
B. cereus	IK1_05630	4TM	None	Unknown	Type I-B	8
Acinetobacter baumannii	AbCdnD	Endonuclease	Cap2, Cap3	Cyclic AMP-AMP-AMP-AMP	Type II-C^{AAAA}	12

most common effector class in type I and type IV CBASS systems, and are rare in type III systems (Fig. 2d).

Several protein families (Pfams) have been designated by Aravind and colleagues[11] to identify CBASS-associated 2TM effector domains (Table 1). A minority of 2TM effectors contain an additional functional domain annotated as a NUDIX hydrolase[11] (Table 1). This domain is known to cleave nucleoside diphosphate molecules linked to other moieties[21] and, therefore, this subclass of effectors probably exerts another function instead of, or in addition to, forming membrane pores.

Some CBASS systems encode effector proteins with four transmembrane helices (4TM). This is possibly a fusion of two 2TM effectors, containing two distinct 2TM effector domains. A 4TM effector was demonstrated to promote infection-mediated cell death in the type I CBASS system of *B. cereus* VD146 (ref. [5]). Interestingly, about 10% of the systems with a 4TM effector have two cyclase genes (compared with ~1% in systems with other effectors), suggesting that the effector may respond to signals from the two cyclases (see below).

Finally, several CBASS effectors (69 of the cases) involve a single transmembrane domain (1TM). Most of these are annotated with a domain of unknown function DUF1043.

Endonuclease (C). Another common effector type encodes an endonuclease domain (Fig. 2d). Endonuclease domains are found mostly in CBASS types II and III. Four different subclasses of endonuclease domains appear with these effectors: HNH endonuclease (CBASS types I and II), NucC endonuclease (mostly in type III CBASS), a recently identified endonuclease of Pfam PF14130 (formerly called DUF4297)[12] and, rarely, Mrr or calcineurin-like nucleases. The NucC-domain effector protein of type III CBASS from *E. coli* MS115-1 was shown to be activated by its cognate cyclic oligonucleotide (cyclic triadenylate) and, once activated, it was shown to degrade double-stranded DNA indiscriminately to fragments sized 50–100 bp (ref. [10]). DNA degradation during the infection process eliminates both phage and host genomes, aborting the infection and leading to cell death[10]. In a similar manner, effector proteins that contain the DUF4297 endonuclease (denoted Cap4) or the HNH endonuclease domain (Cap5) were shown to degrade DNA into short fragments when activated by their cognate cyclic-oligonucleotide molecules[12].

TIR domains (D). Toll/interleukin-1 receptor (TIR) domains were originally found in pathogen-sensing innate-immunity proteins in eukaryotes[22]. These domains were recently shown to participate in a widespread anti-phage defence system called Thoeris, the mechanism of action of which is yet to be deciphered[2]. The presence of TIR domains within CBASS effectors mark TIRs as possibly exerting a cell-killing function. Three subclasses of TIR domains are associated with CBASS systems: (1) a TIR-like domain fused to a STING domain in type I CBASS; (2) a TIR-like domain fused to a cyclic-oligonucleotide-sensor domain annotated as Pfam PF18145 in type II systems; and (3) a TIR-like domain fused to a cyclic-oligonucleotide-sensor domain annotated as Pfam PF18178 in type III systems.

Phosphorylase/nucleosidase superfamily (E). Protein domains with this annotation are found rarely (60 cases in our set) in effector genes from type III CBASS systems.

Peptidase (F). Peptidases of the caspase-like superfamily are found rarely (20 such cases in our set) in effector genes that mostly belong to type I CBASS. We postulate that this domain may be involved in cell suicide, perhaps by cleaving essential cellular proteins.

Classification axis III: the signalling molecule. In a recent comprehensive study, Kranzusch and colleagues showed that the bacterial cGAS protein, which was originally detected in *V. cholerae*[6], is part of a large and diverse family of oligonucleotide cyclases (CD-NTases) that is widespread in microorganisms[8]. Various cyclases from the CD-NTase family have been experimentally shown to generate a variety of cyclic di- and trinucleotides, including cyclic cGAMP, cyclic UMP-AMP, cyclic UMP-UMP and cyclic AMP-AMP-GMP[8]. CBASS systems of the same type can encode different cyclases that produce different signalling molecules (Table 2). We propose to include the biologically active signalling molecule of CBASS systems as a superscript annotation, such that, for example, a type II CBASS with a phospholipase effector that utilizes cyclic UMP-AMP signalling molecule will be named CBASS type II-A^{UA} (Table 2).

A phylogenetic tree of the oligonucleotide cyclase proteins is shown in Extended Data Fig. 1. This tree largely follows the trees for this family of proteins presented in refs. [5,8], and we now add the updated CBASS and effector types overlaid on the tree (Extended Data Fig. 1).

CBASS effector genes must encode cyclic-oligonucleotide-sensing domains that match the signalling molecule produced by their cognate oligonucleotide cyclase. About 30% of the effector proteins in CBASS systems encode a domain called SAVED[11], which was recently shown to form the oligonucleotide-sensing platform in the effector protein. Divergence in the nucleotide-binding pockets of SAVED enables the specific recognition of a large variety of cyclic oligonucleotides that differ in ring size and composition[12]. Other protein domains that are found in CBASS effector proteins that

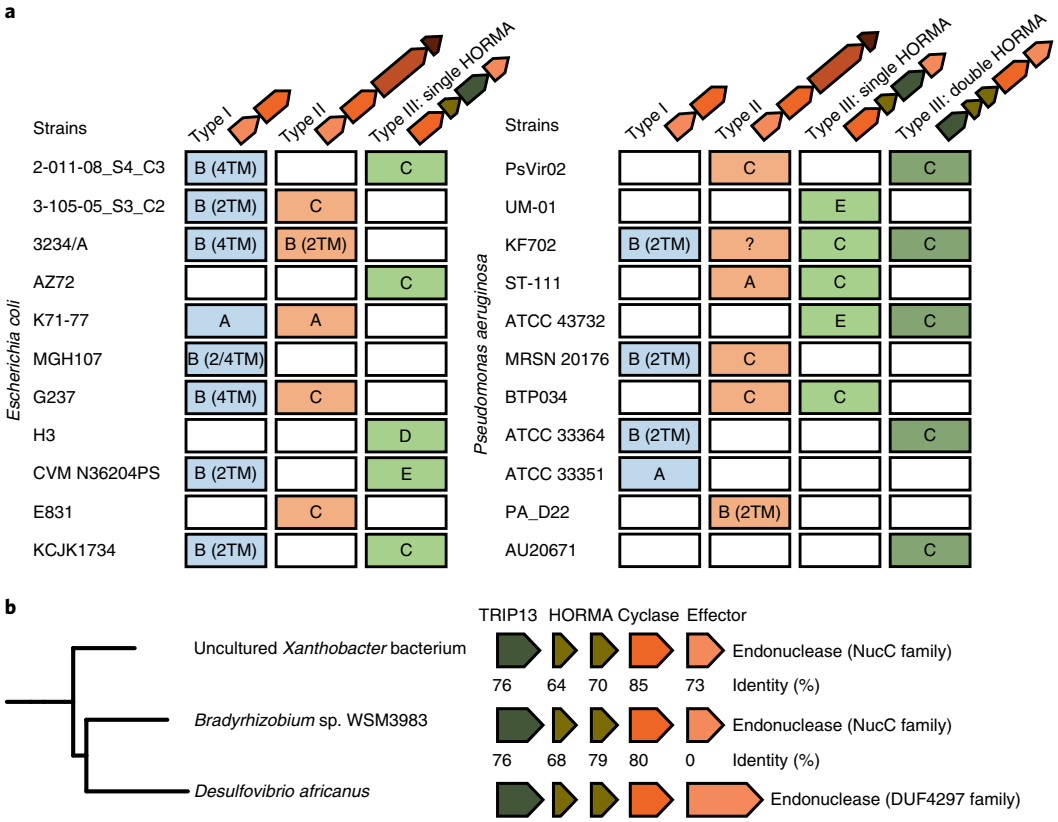


Fig. 3 | Rapid gain and loss of and gene shuffling in CBASS systems. a, The presence of different CBASS systems in closely related genomes. Each row represents a different strain of either *E. coli* (left) or *Pseudomonas aeruginosa* (right), each column corresponds to a different CBASS type. The coloured boxes indicate the presence of the CBASS system. The white boxes depict absence of CBASS. The letter within the box (A–E) represents the effector type. A, phospholipase; B, transmembrane domains; C, endonuclease; D, TIR domain; E, phosphorylase/nucleosidase. ST-111, ST-111 38_London_12_VIM_2_08_12. **b**, Phylogenetic tree of three highly similar cyclases (80–85% protein sequence identity) that belong to a type III CBASS. The percentage identity between each protein and its cognate protein in the neighbouring leaf of the tree is shown.

do not encode SAVED domains are predicted to form additional oligonucleotide-sensing platforms (Table 1).

Variations on a theme. In addition to the common CBASS configurations presented above, several variations are worth mentioning. There are 78 cases in which the CBASS operon contains two oligonucleotide cyclase enzymes (Supplementary Tables 1–5). Interestingly, the majority (53%) of these CBASSs encode a 4TM effector that seems to be, as described above, a fusion of two 2TM effectors. It is possible that these systems represent an actual fusion of two CBASS effectors, and that each half of the effector senses the signalling molecule generated by one of the two associated oligonucleotide cyclases but not the other.

In 914 of the CBASS systems we examined, the oligonucleotide cyclase gene encodes an extra domain (Pfam PF18134). The vast majority (98%) of these cases are type I systems (Supplementary Table 1). It has previously been suggested that this protein domain functions as a cyclic-oligonucleotide sensor[11], but its presence in the oligonucleotide cyclase protein suggests that its role may be to identify the invading phage. The abundance of the PF18134 protein domain specifically in type I CBASS systems may also imply that this domain replaces the function of the ancillary genes, which are naturally missing from type I systems.

An interesting variant of CBASS systems lacks the hallmark oligonucleotide cyclase gene. Instead, these systems contain a gene that is annotated as an adenylate/guanylate cyclase, as well as an effector gene with a cyclic AMP (cAMP)-binding domain instead of the common cyclic-oligonucleotide-binding domain (Supplementary Tables 6 and 7). We found 460 such systems (303 of type I CBASS and 127 of type II CBASS), and it is possible that, in these systems, the signalling molecule is a single nucleotide with a cyclic monophosphate rather than a cyclic oligonucleotide. As cAMP is frequently used for housekeeping signalling functions in bacteria[23], it is probable that the signalling molecule is cGMP. Alternatively, it is possible that the adenylate cyclase gene produces a cAMP variant that differs from the housekeeping cAMP (for example, perhaps the cyclic phosphate is formed by covalent linkage with the 2′ carbon of the ribose instead of the 3′ carbon).

Evolutionary shuffling of CBASS systems and genes. Many anti-phage defence systems were shown to undergo a high level of horizontal gene transfer[1,24,25]. It was suggested that bacteria use horizontal gene transfer to access defence mechanisms encoded by the pan-immune system of closely related strains[1]. In accordance with this, we found that different strains of the same species can host different CBASS types and effectors, sometimes with multiple systems in the same genome (Fig. 3a). This chequered pattern of CBASS distribution in closely related genomes suggests that CBASS systems are rapidly gained and lost, consistent with the present theory on the pan-immune system of bacteria[1]. We found 561 genomes that contain more than one CBASS, with some genomes containing four or even five of such systems (Supplementary Tables 1–7).

Examination of the phylogenetic tree of oligonucleotide cyclases shows evidence for shuffling of cyclase and effector genes between

CBASS systems (Extended Data Fig. 1). Cyclase genes that are phylogenetically close to each other on the tree (that is, those that have similar sequences) can be associated with different effector types and even with different CBASS types (Extended Data Fig. 1). In some cases, we found highly similar systems (>80% identity when comparing the cyclase proteins) that have completely different effectors (Fig. 3b). Presumably, such a shuffling of effector genes between CBASS systems is evolutionarily driven by the necessity to mitigate infection by phages that can inhibit or overcome some of the effectors.

Discussion

We have classified the vast diversity of CBASS systems into a streamlined nomenclature that will hopefully help researchers to give recognizable names to the CBASS systems that they study. The three classification identifiers are operon organization, effector type and signalling molecule, and applying this classification strategy will enable straightforward description of most CBASSs, providing a common language for research in the field. The complexity of CBASS systems parallels that of CRISPR systems, for which similar past classifications were instrumental in generating a common language for research in the field[26,27].

CBASS types differ by the content of their ancillary genes. In type II and type III CBASS systems, the proteins encoded by the ancillary genes have counteractive functions. In type III CBASS, the Cap7 HORMA-domain protein activates the oligonucleotide cyclase by directly binding to it, and the Cap6 TRIP13 protein inhibits this activation by dissociating Cap7 from the cyclase enzyme[9]. In a similar manner, it is predicted that the Cap2 ancillary protein of type II CBASS systems uses its E1 and E2 domains to conjugate a (yet unidentified) moiety to a target protein to possibly activate the cyclase enzyme, while the Cap3 JAB-domain isopeptidase protein would counteract this activity by removing this moiety from its target.

An interesting parallel of the CBASS system is found in type III CRISPR–Cas systems. After recognising foreign nucleic acids through the crRNA–protein complex of type III CRISPR–Cas, the Cas10 subunit of the complex generates a cyclic oligoadenylate molecule (comprising 2–6 adenosine monophosphate molecules conjugated in a cyclic form)[28,29]. This molecule then binds to and activates the RNase effector Csm6 that indiscriminately degrades the RNA of both host and phage, presumably leading to cell dormancy or death[28,29]. In at least some cases, bona fide CBASS effectors, such as the NucC endonuclease, function as effectors of type III CRISPR–Cas systems[10], suggesting that these two defence systems share functional components.

Interactions between bacteria and phages have led to the evolution of multiple and diverse bacterial immune systems. Many of these systems, including CBASS, are absent from many model organisms and have only been discovered very recently by studying the genomes of non-model organisms[1,2,5,30–33]. It is probable that many more novel defence systems will be discovered that will showcase the manifold ways in which bacteria have evolved to defend against virus infection.

Methods

Identification of CBASS cyclases and operon types. Protein sequences of all of the genes in 38,167 bacterial and archaeal genomes were downloaded from the Integrated Microbial Genomes (IMG) database[34] in October 2017. These proteins were clustered using the 'cluster' option of MMseqs2 (release 2-1c7a89; ref. [35]), with the default parameters. The sequences of each cluster were aligned using Clustal Omega[36]. Multiple sequence alignments of clusters of 10 or more proteins were scanned with HHsearch[37] using the 50% gap rule (-M 50) against the PDB_mmCIF70 (ref. [38]) and Pfam32 (ref. [39]) databases. For clusters with less than ten proteins, we used a step of HHblits[37] with two iterations before scanning with HHsearch. Clusters with HHsearch hits to one of the cGAS entries (Protein Data Bank (PDB) codes: 4LEV, 4MKP, 4O67, 5VDR, 5V8H, 4LEW, 5VDP, 4KM5, 4O68, 4O69, 5V8J, 5V8N, 5V8O, 5VDO, 5VDQ, 5VDS, 5VDT, 5VDU, 5VDV, 5VDW, 4XJ5, 4XJ1, 4XJ6, 4XJ3 and 4XJ4; and Pfam: PF03281) with >90% probability in the top 30 hits were taken for manual analysis. Fragmented cyclase genes were counted as a single gene in further analyses.

To identify the CBASS operon types, the genomic environments spanning ten genes upstream and downstream of each of the predicted oligonucleotide cyclase genes were searched to identify conserved gene cassettes, as previously described[2,5]. Predicted systems were manually reviewed and unrelated genes (such as mobilome genes and genes of other defence systems) were omitted.

To search for unannotated JAB proteins, the nucleotide sequences of E1–E2 genes from type II CBASS operons that do not contain JAB were extracted, along with the 2,000 nucleotides upstream and downstream. These sequences were searched using BLASTx against the JAB protein from the CBASS operon of *E. coli* TW11681 (ref. [5]). Hits with an *e* value of <0.01 were considered to be cases in which JAB was present. The same process was used to detect unannotated E2 genes.

Phylogenetic analysis. Genomes that contain CBASS were counted for each phylum. A species was counted as containing CBASS (Fig. 2b) if at least one of the genomes belonging to the species had a CBASS system.

To generate the phylogenetic tree in Extended Data Fig. 1, cyclases that did not have ten genes upstream and downstream were removed. The 'clusthash' option of MMseqs2 (release 6-f5a1c) was used to remove protein redundancies (using the '--min-seq-id 0.9' parameter). Sequences that were shorter than 200 amino acids were also removed. The human cGAS protein (UniProt: Q8N884) was added, as well as the human oligoadenylate synthase genes (UniProt: P00973, P29728 and Q9Y6K5); these were used as an outgroup. Sequences were aligned using clustal-omega v.1.2.4 with the default parameters. FastTree[40] was used to generate a tree from the multiple-sequence alignment using the default parameters. iTOL[41] was used for tree visualization. Clades were coloured according to refs. [5,8].

Search for CBASS AMP/GMP cyclases. The genomic environments spanning ten genes upstream and downstream of E1–E2 genes were searched to identify conserved gene cassettes, as previously described[2,5]. Genes associated with E1–E2-containng operons that did not contain an oligonucleotide cyclase were mapped to their respective cluster (see the explanation for the protein clustering above). Multiple-sequence alignments of proteins in these clusters were scanned with HHsearch[37] using the 50% gap rule (-M 50) against the Pfam32 (ref. [39]) database. Clusters with hits to Pfam PF00211 with >90% probability in the top 30 hits were then manually examined.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All genomic data that support the findings of this study are available at IMG (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi). Accession codes for all data are provided in Supplementary Tables 1–7. PDB and Pfam databases are available at the HHsuite database page (http://wwwuser.gwdg.de/~compbiol/data/hhsuite/databases/hhsuite_dbs/).

Received: 19 March 2020; Accepted: 16 July 2020;
Published online: 24 August 2020

References

1. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
2. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
3. Rostøl, J. T. & Marraffini, L. (Ph)ighting phages: how bacteria resist their parasites. *Cell Host Microbe* **25**, 184–194 (2019).
4. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
5. Cohen, D. et al. Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).
6. Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**, 358–370 (2012).
7. Severin, G. B. et al. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **115**, E6048–E6055 (2018).
8. Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
9. Ye, Q. et al. HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. *Mol. Cell* **77**, 709–722 (2020).
10. Lau, R. K. et al. Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Mol. Cell* **77**, 723–733 (2020).

11. Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015).
12. Lowey, B. et al. CBASS immunity uses CARF-related effectors to sense 3′–5′- and 2′–5′-linked cyclic oligonucleotide signals and protect bacteria from phage infection. *Cell* **182**, 38–49.e17 (2020).
13. Aravind, L. & Koonin, E. V. The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair. *Trends Biochem. Sci.* **23**, 284–286 (1998).
14. Rosenberg, S. C. & Corbett, K. D. The multifaceted roles of the HORMA domain in cellular signaling. *J. Cell Biol.* **211**, 745–755 (2015).
15. Vader, G. Pch2(TRIP13): controlling cell division through regulation of HORMA domains. *Chromosoma* **124**, 333–339 (2015).
16. Reader, J. S., Metzgar, D., Schimmel, P. & de Crécy-Lagard, V. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J. Biol. Chem.* **279**, 6280–6285 (2004).
17. McCarty, R. M., Somogyi, A., Lin, G., Jacobsen, N. E. & Bandarian, V. The deazapurine biosynthetic pathway revealed: in vitro enzymatic synthesis of PreQ(0) from guanosine 5′-triphosphate in four steps. *Biochemistry* **48**, 3847–3852 (2009).
18. Okada, N. et al. Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction. *J. Biol. Chem.* **254**, 3067–3073 (1979).
19. Thiaville, J. J. et al. Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc. Natl Acad. Sci. USA* **113**, E1452–E1459 (2016).
20. David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950 (2007).
21. McLennan, A. G. The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.* **63**, 123–143 (2006).
22. Akira, S. & Takeda, K. Toll-like receptor signalling. *Nat. Rev. Immunol.* **4**, 499–511 (2004).
23. Botsford, J. L. & Harman, J. G. Cyclic AMP in prokaryotes. *Microbiol. Rev.* **56**, 100–122 (1992).
24. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary genomics of defense systems in archaea and bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
25. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
26. Makarova, K. S. et al. Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).
27. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
28. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609 (2017).
29. Niewoehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).
30. Ofir, G. et al. DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).
31. Goldfarb, T. et al. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
32. Ofir, G. & Sorek, R. Contemporary phage biology: from classic models to new insights. *Cell* **172**, 1260–1270 (2018).
33. Lopatina, A., Tal, N. & Sorek, R. Abortive infection: bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.* **7**, 1 (2020).
34. Chen, I. M. A. et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019).
35. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
36. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
37. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform.* **20**, 473 (2019).
38. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
39. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
40. Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
41. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).

Acknowledgements

We thank the members of the Sorek laboratory for comments on earlier versions of this manuscript. A.M. was supported by a fellowship from the Ariane de Rothschild Women Doctoral Program and, in part, by the Israeli Council for Higher Education via the Weizmann Data Science Research Center. R.S. was supported in part by the Israel Science Foundation (personal grant no. 1360/16), the European Research Council (grant no. ERC-CoG 681203), the German Research Council (DFG) priority program SPP 2002 (grant no. SO 1611/1-1), the Israeli Council for Higher Education through the Weizmann Data Science Research Center, the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, the Minerva Foundation with funding from the Federal German Ministry for Education and Research and the Knell Family Center for Microbiology.

Author contributions

A.M. collected and analysed the data and wrote the paper. S.M. and G.A. were involved in the classification of CBASS systems. R.S. supervised the study and wrote the paper.

Competing interests

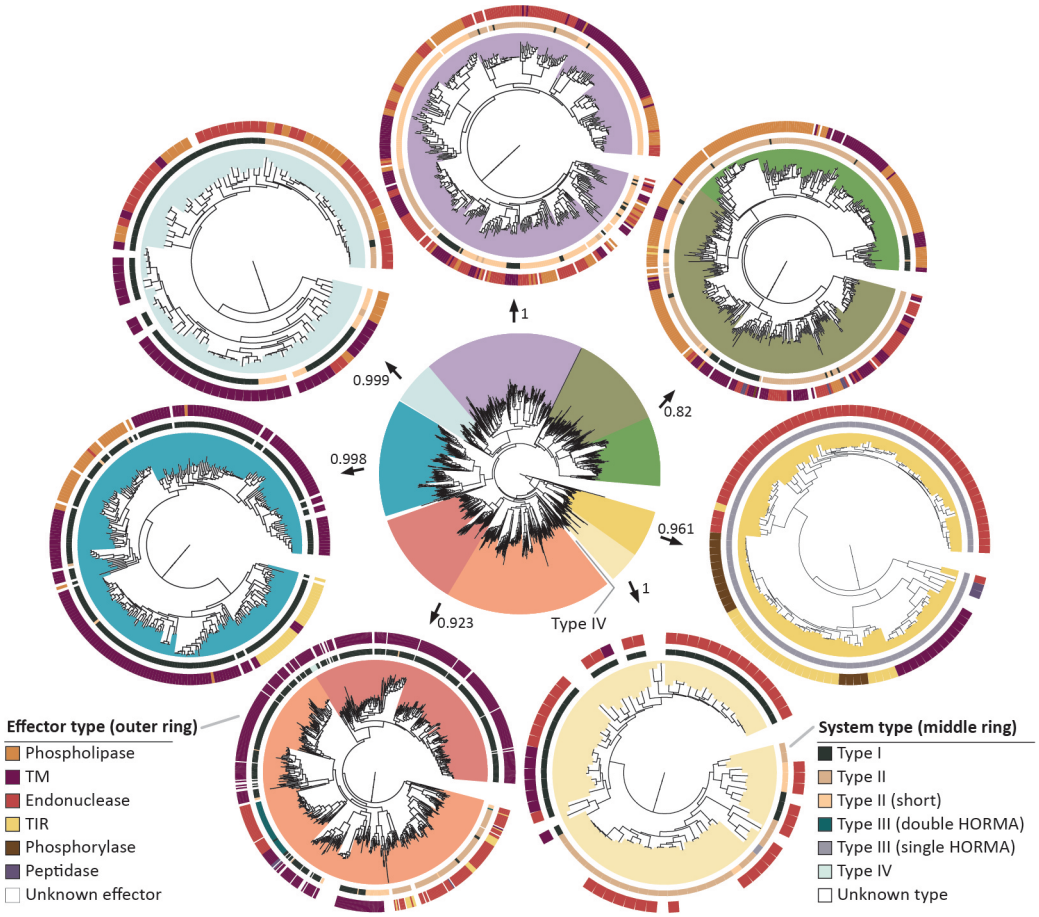
R.S. is a scientific cofounder and consultant of BiomX, Pantheon Bioscience and Ecophage.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-020-0777-y.

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0777-y.

Correspondence and requests for materials should be addressed to R.S.



Extended Data Fig. 1 | Phylogenetic analysis of oligonucleotide cyclases (CD-NTase) and their CBASS types. The phylogenetic tree of all cyclases, as depicted and colored in refs (5 and 8) is presented in the center. Each clade is then expanded and presented in the periphery as a circular tree to increase resolution. Outer ring depicts the effector type; middle ring depicts the system type. Numbers next to each clade represent the bootstrap value for that node in the central tree.

nature research

Corresponding author(s): Rotem Sorek

Last updated by author(s): Jul 6, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect the data.
Data analysis	MMseqs2 (2-1c7a89), Clustal Omega (1.2.4), HHsuite (3.2.0), BLAST (2.5.0), FastTree (2.1.10), iTOL (5.5.1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All genomic data that support the findings of this study are available in IMG, https://img.jgi.doe.gov/cgi-bin/mer/main.cgi. Accession codes for all data are provided in Supplementary tables 1 and 2.
PDB and Pfam databases are available at http://wwwuser.gwdg.de/~compbiol/data/hhsuite/databases/hhsuite_dbs/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size is the full set of bacterial and archaeal genomes available for download from IMG on October 2017.
Data exclusions	To generate the phylogenetic tree in Figure 3, cyclases that do not have 10 genes upstream and downstream were removed. The 'clusthash' option of MMseqs2 (release 6-f5a1c) was used to remove protein redundancies (using the '–min-seq-id 0.9' parameter). Sequences shorter than 200 amino acids were also removed. These exclusion criteria were pre-established
Replication	N/A
Randomization	All data was analyzed no randomization needed.
Blinding	Blinding was irrelevant, all data was analyzed systematically.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging